

Generalized I-MMSE for K -User Gaussian Channels

Samah A. M. Ghanem, *Senior Member, IEEE*

Abstract—In this paper, we generalize the fundamental relation between the mutual information and the minimum mean squared error (MMSE) by Guo, Shamai, and Verdú [1] to K -User Gaussian channels. We prove that the derivative of the multiuser mutual information with respect to the signal to noise ratio (SNR) is equal to the total MMSE plus a covariance term with respect to the cross correlation of the multiuser input estimates, the channels and the precoding matrices. We shed light that such relation is a generalized I-MMSE with one step lookahead and lookback, applied to the Successive Interference Cancellation (SIC) in the decoding process.

I. INTRODUCTION

Duncan, in [2] showed that for the continuous-time additive white Gaussian noise (AWGN) channel, the filtering minimum mean squared error is twice the input output mutual information for any underlying signal distribution. This has illuminated intimate connections between information theory and estimation theory which has been emphasized by Guo, Shamai, and Verdú in a seminal paper [1]. More specifically, Guo et al. have shown that in the classical problem of information transmission through the conventional AWGN channel, the derivative of the mutual information with respect to the SNR is equal to the smoothing minimum mean squared error; a relationship that holds for scalar, vector, discrete-time and continuous-time channels regardless of the input statistics. The relevance of these recent connections comes from the fact that mutual information and MMSE are two canonical operational measures in information theory and estimation theory. Later Palomar and Verdú extended this relation to linear vector Gaussian channels [3], [4]. The mutual information was also represented as an integral of a certain measure of the estimation error in Poisson channels [5], [6]. Most recently, Ghanem in [7], [8], derived the gradient of the joint, conditional and non-conditional mutual information with respect to arbitrary parameters of the multiple access Gaussian channels, a relation that extends the case of mutually interfering inputs in linear vector Gaussian channels to the case of multiple non-mutually interfering inputs and with mutual interference, a starting point to the results in this work. The results of this work extends and deepens the two-user case simple framework presented in [9] to a K -user framework with a comprehensive interpretation to the interference and the successive interference cancellation in the decoding process. The implications of a framework involving key quantities in information theory and estimation theory are countless both from the theoretical [10], [11] and the more practical perspective, see e.g. [12], [13], [14], [15]. For instance, most recently, Ghanem in [16] proved the existence of such connections on network-coded flows

over noisy networks, opening a new horizon of engineering design of precoding and decoding solutions adapted to network topology awareness. Furthermore, connections between information measures and estimation measures allow for finding explicit closed form expressions of the mutual information for binary inputs, particularly ones for BPSK and QPSK over the single input single output (SISO) channel, [1], [17], [18]. The author in [19] provides a new look - that relies on the user decoding order - towards finding such closed forms for the multiuser case. Therefore, it is of particular importance to address connections between information theory and estimation theory for the multiuser (K -users) case. First, it allows for interpreting the interference along the SIC decoding process of users. Second, it goes aligned with and beyond the communications applications and with precise analytical lookahead and lookback framework in mixtures of processes within its application to the communications framework.

In this paper, we first revisit the connections between the mutual information and the MMSE for the K -users Gaussian channels, see also [8], [15]. Therefore, the fundamental relation between the derivative of the mutual information and the MMSE, known as I-MMSE identity, and defined for point to point channels with any noise or input distributions in [1] is not anymore suitable for the multiuser case. Therefore, we generalize the I-MMSE relation to the multiuser case, interested reader can refer to the extended version in [19].

Throughout the paper, the following notation is employed, boldface uppercase letters denote matrices, lowercase letters denote scalars. The superscript, $(\cdot)^{-1}$, $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^\dagger$ denote the inverse, transpose, conjugate, and conjugate transpose operations. The $\mathbb{E}[\cdot]$ denotes the expectation operator. The $\|\cdot\|$ and $\text{Tr}\{\cdot\}$ denote the Euclidean norm, and the trace of a matrix, respectively.

The rest of the paper is organized as follows, Section II introduces the system model. Section III introduces the new fundamental relation between the multiuser mutual information and the MMSE. Section IV provides the conditional and non-conditional components of the K -user I-MMSE identity, where a precise SIC decoding process defines per-user components in information measures and estimation measures. Then, we highlight how a generalized I-MMSE is one with lookahead and lookback components, given the application provided. Section V concludes the paper.

II. SYSTEM MODEL

Consider the deterministic complex-valued vector channel,

$$\mathbf{y} = \sum_{k=1}^K \sqrt{\text{snr}} \mathbf{H}_k \mathbf{P}_k \mathbf{x}_k + \mathbf{n}, \quad (1)$$

where the $n_r \times 1$ dimensional vector \mathbf{y} and the $n_t \times 1$ dimensional vectors $\mathbf{x}_1, \dots, \mathbf{x}_K$ represent, respectively, the received vector and the independent zero-mean unit-variance transmitted information vectors from each user $k = 1, \dots, K$ input to the multiuser channel. The distributions of the inputs are not fixed, not necessarily Gaussian nor identical. The $n_r \times n_t$ complex-valued matrices $\mathbf{H}_1, \dots, \mathbf{H}_K$ correspond to the deterministic channel gains for the K -users input channels (known to both encoder and decoder) and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the $n_r \times 1$ dimensional complex Gaussian noise with independent zero-mean unit-variance components. The $n_t \times n_t$ $\mathbf{P}_1, \dots, \mathbf{P}_K$ are precoding matrices that do not increase the transmitted power¹. For such a system, the conditional probability distribution of the Gaussian noise is defined as:

$$p_{\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K) = \frac{1}{\pi^{n_r}} e^{-\|\mathbf{y} - \sum_{k=1}^K \sqrt{\text{snr}} \mathbf{H}_k \mathbf{P}_k \mathbf{x}_k\|^2} \quad (2)$$

Additionally, the probability density function for the received vector \mathbf{y} is defined as:

$$p_{\mathbf{y}}(\mathbf{y}) = \sum_{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K} p_{\mathbf{y}|\mathbf{x}_1, \dots, \mathbf{x}_K}(\mathbf{y}|\mathbf{x}_1, \dots, \mathbf{x}_K) \prod_{k=1}^K p_{\mathbf{x}_k}(\mathbf{x}_k). \quad (3)$$

Capitalizing on the previous fundamental definitions, we present the main contributions in the next sections.

III. GENERALIZED K -USER I-MMSE

The first contribution is given in the following theorem, which provides a generalization of the I-MMSE identity to the K -users case.

Theorem 1: The relation between the derivative of the multiuser mutual information with respect to the SNR and the non-linear MMSE for a multiuser Gaussian channel satisfies:

$$\frac{dI(\text{snr})}{d\text{snr}} = \text{mmse}(\text{snr}) + \psi(\text{snr}) \quad (4)$$

Where,

$$\text{mmse}(\text{snr}) = \sum_{k=1}^K \text{Tr} \{ \mathbf{H}_k \mathbf{P}_k \mathbf{E}_k (\mathbf{H}_k \mathbf{P}_k)^\dagger \} \quad (5)$$

$$\psi(\text{snr}) = - \sum_{k=1}^K \sum_{j=1, j \neq k}^K \text{Tr} \{ \mathbf{H}_k \mathbf{P}_k \mathbb{E}_{\mathbf{y}} [\mathbb{E}_{\mathbf{x}_k|\mathbf{y}} [\mathbf{x}_k|\mathbf{y}] \mathbb{E}_{\mathbf{x}_j|\mathbf{y}} [\mathbf{x}_j|\mathbf{y}]^\dagger] \times (\mathbf{H}_j \mathbf{P}_j)^\dagger \}$$

Proof: The proof is provided in [19]. ■

¹Notice that $\mathbf{H}_k \mathbf{P}_k$ matrix can correspond to any measurement matrix \mathbf{M}_k of a system of matched mixtures of arbitrary processes contaminated by AWGN.

The per-user MMSE is given as follows:

$$\mathbf{E}_k = \mathbb{E}_{\mathbf{y}} [(\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^\dagger] \quad (6)$$

The non-linear input estimates of each user input is given as follows:

$$\begin{aligned} \hat{\mathbf{x}}_k &= \mathbb{E}_{\mathbf{x}_k|\mathbf{y}} [\mathbf{x}_k|\mathbf{y}] \\ &= \sum_{\mathbf{x}_1, \dots, \mathbf{x}_K} \frac{\mathbf{x}_k p_{\mathbf{y}|\mathbf{x}_1, \dots, \mathbf{x}_K}(\mathbf{y}|\mathbf{x}_1, \dots, \mathbf{x}_K) \prod_{k=1}^K p_{\mathbf{x}_k}(\mathbf{x}_k)}{p_{\mathbf{y}}(\mathbf{y})} \end{aligned} \quad (7)$$

Note that the term $\text{mmse}(\text{snr})$ is due to the K -users MMSEs, particularly,

$$\begin{aligned} \text{mmse}(\text{snr}) &= \text{mmse}_1(\text{snr}) + \text{mmse}_2(\text{snr}) \\ &+ \dots + \text{mmse}_{K-1}(\text{snr}) + \text{mmse}_K(\text{snr}) \end{aligned} \quad (8)$$

and $\psi(\text{snr})$ are covariance terms that appear due to the covariance of the K -users interference. In particular, this term corresponds to the interference of $K - 1$ users to one user, and how each user contribute with an interference term to the other $K - 1$ users. Those terms are with respect to the channels, precoders, and non-linear estimates of the user inputs as will be precisely provided in the next section.

When the covariance terms vanish to zero, the derivative of the mutual information with respect to the SNR will be equal to the MMSE with respect to the SNR. This applies to the relation for point to point communications or in other words if no interference is encountered by any user. Therefore, the result of Theorem 1 is a generalization of such connection between the two canonical operational measures in information theory and estimation theory - the mutual information and the MMSE - and boils down to the result of Guo et. al, [1] under certain conditions which are: (i) when the cross correlation between the inputs estimates equals zero (ii) when interference can be neglected. Therefore, when the term $\psi(\text{snr})$ equals zero. The derivative of the mutual information with respect to the SNR equals the total $\text{mmse}(\text{snr})$:

$$\frac{dI(\text{snr})}{d\text{snr}} = \text{mmse}(\text{snr}) \quad (9)$$

which matches the result by Guo et. al in [1].

Such generalized fundamental relation between the change in the multiuser mutual information and the SNR is of particular relevance. Firstly, it allows understanding the behavior of per-user rates with respect to the interference due to the mutual interference and the interference of other $K - 1$ users in terms of their power levels and channel strengths. In addition, the result allows us to be able to quantify the losses incurred due to the interference in terms of bits. Moreover, this new relation yet captures the mutual interference introduced if cooperation is considered among nodes. Thus, this generalization is generic for point-to-point, and linear vector Gaussian channels. Extended version can be found in [19].

IV. K -USER CONDITIONAL AND NON-CONDITIONAL I-MMSE WITH SIC DECODING

We capitalize on the new fundamental relation to extend the derivative with respect to the SNR to K -user conditional and non-conditional mutual information in the SIC process. Capitalizing on the chain rule of the mutual information, the joint mutual information of K -users is given as follows:

$$\begin{aligned} I(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K; \mathbf{y}) &= I(\mathbf{x}_1; \mathbf{y}) + I(\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_K; \mathbf{y} | \mathbf{x}_1) \\ &= I(\mathbf{x}_1; \mathbf{y}) + I(\mathbf{x}_2; \mathbf{y} | \mathbf{x}_1) + I(\mathbf{x}_3, \dots, \mathbf{x}_K; \mathbf{y} | \mathbf{x}_1, \mathbf{x}_2) \\ &= I(\mathbf{x}_1; \mathbf{y}) + I(\mathbf{x}_2; \mathbf{y} | \mathbf{x}_1) + I(\mathbf{x}_3; \mathbf{y} | \mathbf{x}_1, \mathbf{x}_2) \\ &\quad + \dots + I(\mathbf{x}_K; \mathbf{y} | \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{K-1}) \end{aligned} \quad (10)$$

Therefore, through this observation we can conclude the following theorem.

Theorem 2: The relation between the derivative of the K -user conditional and the non-conditional mutual information and their corresponding minimum mean squared error, for a step by step SIC decoding, satisfies, respectively:

$$\frac{dI(\mathbf{x}_1; \mathbf{y})}{dsnr} = mmse_1(\gamma snr) \quad (11)$$

$$\frac{dI(\mathbf{x}_2; \mathbf{y} | \mathbf{x}_1)}{dsnr} = mmse_2(\lambda snr) + \psi_{1,2}(snr) + \psi_{2,1}(snr) \quad (12)$$

...

$$\frac{dI(\mathbf{x}_k; \mathbf{y} | \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{K-1})}{dsnr} = mmse_K(snr) + \psi_{1,K}(snr) + \psi_{K,1}(snr) \quad (13)$$

where,

$$\begin{aligned} \psi_{k,1}(snr) &= - \sum_{\ell=k}^1 Tr\{\mathbf{H}_k \mathbf{P}_k \mathbb{E}_{\mathbf{y}}[\mathbb{E}_{\mathbf{x}_k|\mathbf{y}}[\mathbf{x}_k|\mathbf{y}]\mathbb{E}_{\mathbf{x}_\ell|\mathbf{y}}[\mathbf{x}_\ell|\mathbf{y}]^\dagger] \times \\ &\quad (\mathbf{H}_\ell \mathbf{P}_\ell)^\dagger\} \end{aligned}$$

and,

$$\begin{aligned} \psi_{1,k}(snr) &= - \sum_{\ell=1}^k Tr\{\mathbf{H}_k \mathbf{P}_k \mathbb{E}_{\mathbf{y}}[\mathbb{E}_{\mathbf{x}_\ell|\mathbf{y}}[\mathbf{x}_\ell|\mathbf{y}]\mathbb{E}_{\mathbf{x}_k|\mathbf{y}}[\mathbf{x}_k|\mathbf{y}]^\dagger] \times \\ &\quad (\mathbf{H}_k \mathbf{P}_k)^\dagger\} \end{aligned}$$

Proof: Details of the proof follows from Theorem 1. Then, taking the derivative of both sides of (10), and subtracting the derivative of $I(\mathbf{x}_1; \mathbf{y})$ which is equal to user 1 $mmse_1(\gamma snr)$, γ is a scaling factor, due to the fact that \mathbf{x}_1 is decoded first considering the other $K-1$ users' inputs as noise. Then, subtracting the derivative of $I(\mathbf{x}_2; \mathbf{y} | \mathbf{x}_1)$ which corresponds to user 2 $mmse_2(\lambda snr)$, λ is a scaling factor, due to the fact that \mathbf{x}_2 is decoded second considering the other $K-2$ users' (except user 1) inputs as noise and based on the knowledge of \mathbf{x}_1 , we have the added covariance $\psi_{1,2}(snr) + \psi_{2,1}(snr)$, where $\psi_{1,2}$ appears as interpretation of the user 2 interference on user 1, and $\psi_{2,1}(snr)$ appears as interpretation of the first user interference on user 2. Repeating the same steps, until user K , we have the derivative of $I(\mathbf{x}_k; \mathbf{y} | \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{K-1})$

equals to a non-scaled $mmse_K(snr)$ plus the covariance $\psi_{1,2,\dots,K-2,K}(snr) + \psi_{K,K-1,K-2,\dots,1}(snr)$ caused by the K -th user interference on the other $K-1$ users and the $K-1$ users interference on the K -th user respectively. And $\psi(snr) = \sum_{\forall\{k,\ell\} \in \{1,\dots,K\}} \psi_{k,\ell}(snr)$ which corresponds to $K!$ covariance terms. Therefore, Theorem 2 has been proved, and matches with Theorem 1 where the sum of the derivatives of the per users' mutual information equals the derivative of the joint mutual information. ■

Of particular relevance is the implication of the derived relations on understanding the achievable rates of interference channels. In particular, such relation allows for better understanding of the changes in the rates due the interferer which is either decoded first or considered as noise in a SIC form. Additionally, the theorem can be analytically interpreted in a different way if we consider that part of the users consider partial set of the users interference as noise. Moreover, if a joint decoding process is in place, the interpretation again changes, just by moving some covariance terms systematically and removing the scaling of the MMSE.

A. The Generalized I-MMSE is an I-IMMSE

In this subsection, we shed light on the impact of such generalized multiuser I-MMSE into a general class of I-MMSE relations with finite lookahead and lookback. In particular, such generalization explains in explicit steps how a one step ahead in the knowledge process via conditioning provides less error, thus more information rates of a user decoded one step ahead. Besides, looking back one step into the successive decoding or interference cancellation process, the relation provides explicit closed forms of the losses encountered due to being blind about other aspects of the process. Therefore, the scaling with larger variance makes the user away from the ultimate knowledge of the process. In turn, such generalization goes beyond the known I-MMSE by Guo, Shamai, Verdú [1] into a general relation that provides direct and clear connection to the I-IMMSE with lookahead and lookback, [20].

In light of (56) in [20] of the lookahead MMSE, the IMMSE can be decomposed to its components with respect to steps i of lookback $-d$ and lookahead d ,

$$lmmse\left(id, \frac{snr}{\gamma_i}\right) = mmse\left(\frac{snr}{\gamma_i}\right) + \int_{-\infty}^{-id} h^2(t) dt \quad (14)$$

where $h(t)$ denotes a conditional variance in the process or an impulse response of a filter, and $i = 1, \dots, \infty$, and $\gamma_1 > \gamma_2 > \gamma_3 > \dots > \gamma_\infty$, where $\gamma_\infty = 1$.

Finally, it is worth to note that the SIC process provides an application that clarifies the difference in the lookback $-d$ and lookahead d MMSE and their corresponding losses and gains in the information rates due to scaling of the SNR and the change in the cross correlation between input estimates. Thus, the information measure connected to looking back or looking forward is not equal, i.e., $I(snr; \mathbf{y}_0^{-d} | \mathbf{y}_\infty^0) \neq I(snr; \mathbf{y}_0^d | \mathbf{y}_{-\infty}^0)$, respectively. Where $\mathbf{y}_{t_i}^{t_f}$ corresponds to an observation window $[t_i - t_f]$ in the time of process \mathbf{y} . The characterization of each is beyond the scope of this work.

V. CONCLUSIONS

We generalize the fundamental relation between the mutual information and the MMSE - I-MMSE identity - to a K -User case decoded via SIC. Such generalization provides a unique application to I-MMSE with lookahead and lookback via quantifying the rates obtained or lost looking ahead or back into the successive decoding process.

VI. ACKNOWLEDGEMENT

The author would like to thank Prof. Muriel Medard for her valuable feedback about the work.

REFERENCES

- [1] D. Guo and S. Shamai and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, pp. 1261–1282, April 2005.
- [2] T. E. Duncan, "On the calculation of mutual information," *SIAM Journal on Applied Mathematics*, vol. 19, no. 1, pp. 215–220, July 1970.
- [3] D. P. Palomar and S. Verdu, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Transactions on Information Theory*, vol. 52, No.1, pp.141-154, January 2006.
- [4] —, "Representation of mutual information via input estimates," *IEEE Transactions on Information Theory*, vol. 53, no. 2, pp. 453–470, 2007.
- [5] D. Guo, S. Verdu, and S. Shamai, "Mutual information and conditional mean estimation in poisson channels," in *IEEE Information Theory Workshop (ITW)*, Oct. 2004, pp. 265–270.
- [6] S. A. M. Ghanem and M. Ara, "The mac poisson channel: Capacity and optimal power allocation," *IAENG Transactions on Engineering Technologies, Lecture Notes in Electrical Engineering*, ©Springer, vol. 170, pp. 45–60, 2013.
- [7] S. A. M. Ghanem, "MAC Gaussian channels with arbitrary inputs: Optimal precoding and power allocation," in *IEEE International Conference on Wireless Communications and Signal Processing (WCSP), Huangshan, China*, October 2012.
- [8] —, "Multiple access Gaussian channels with arbitrary inputs: Optimal precoding and power allocation," *available on Arxiv.org*, 2014.
- [9] S. A. M. Ghanem, "Mutual information and minimum mean-square error in multiuser Gaussian channels," *IEEE MMTC E-Letters, Special issue on Latest advances on Wireless and Mobile Communications and Networking*, 2015, 02 2015.
- [10] S. Verdu and D. Guo, "A simple proof of the entropy-power inequality," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2165–2166, 2006.
- [11] A. Tulino and S. Verdu, "Monotonic decrease of the non-Gaussianity of the sum of independent random variables: A simple proof," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4295–4297, 2006.
- [12] A. Lozano, A. Tulino, and S. Verdu, "Mercury/waterfilling: Optimum power allocation with arbitrary input constellations," *International Symposium on Information Theory (ISIT)*, pp. 1773–1777, 2005.
- [13] —, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 3033–3051, 2006.
- [14] S. A. M. Ghanem, "Optimal power allocation and optimal precoding with multi-cell processing," *IEEE 77th Vehicular Technology Conference: VTC-Spring, Dresden, Germany*, June 2013.
- [15] —, "Multi-cell processing with limited cooperation: A novel framework to timely designs and reduced csi feedback with general inputs," *available on Arxiv.org*, 2014.
- [16] —, "Network coding: Connections between information theory and estimation theory," pp. 1–6, June 2016.
- [17] —, "Mutual information for generalized arbitrary binary input constellations," in *MAP-Tele Workshop*, February 2010.
- [18] —, *Analysis, Modeling, Design, and Optimization of Future Communications Systems: From Theory to Practice*, Ph.D. Thesis, December 2013.
- [19] —, "Multiuser I-MMSE," *available on Arxiv.org*, 2015.
- [20] K. Venkat, T. Weissman, Y. Carmon, and S. Shamai, "Information, estimation, and lookahead in the gaussian channel," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3605–3618, July 2016.